

Robust fitting of Zernike polynomials to noisy point clouds defined over connected domains of arbitrary shape

Diego Rodríguez Ibañez,¹ José A. Gómez-Pedrero,^{2,*} Jose Alonso,^{1,2} and Juan A. Quiroga^{1,3}

¹Indizen Optical Technologies, C/ Santa Engracia 6, 28010 Madrid, Spain

²Applied Optics Complutense Group, Universidad Complutense de Madrid, Faculty of Optics and Optometry, Avda/ Arcos de Jalón 118, 28037 Madrid, Spain

³Applied Optics Complutense Group, Universidad Complutense de Madrid, Faculty of Physics, Ciudad Universitaria s/n, 28040 Madrid, Spain
jagomezp@ucm.es

Abstract: A new method for fitting a series of Zernike polynomials to point clouds defined over connected domains of arbitrary shape defined within the unit circle is presented in this work. The method is based on the application of machine learning fitting techniques by constructing an extended training set in order to ensure the smooth variation of local curvature over the whole domain. Therefore this technique is best suited for fitting points corresponding to ophthalmic lenses surfaces, particularly progressive power ones, in non-regular domains. We have tested our method by fitting numerical and real surfaces reaching an accuracy of 1 micron in elevation and 0.1 D in local curvature in agreement with the customary tolerances in the ophthalmic manufacturing industry.

©2016 Optical Society of America

OCIS codes: (240.6700) Surfaces; (220.1250) Aspherics; (170.4460) Ophthalmic optics and devices; (150.1135) Algorithms.

References and links

1. W. Wang, H. Pottmann, and Y. Liu, "Fitting B-spline curves to point clouds by curvature-based squared distance minimization," *ACM Trans. Graph.* **25**(2), 214–238 (2006).
2. S. Flöry, "Fitting curves and surfaces to point clouds in the presence of obstacles," *Comput. Aided Geom. Des.* **26**(2), 192–202 (2009).
3. F. Remondino, "From point cloud to surface: the modelling and visualization problem," in *International Workshop on Visualization and Animation of Reality-Based 3D Models*, <http://dx.doi.org/10.3929/ethz-a-004655782> (2003).
4. J. G. Hayes and J. Halliday, "The least-squares fitting of cubic spline surfaces to general data sets," *IMA J. Appl. Math.* **14**(1), 89–103 (1974).
5. P. Dierckx, *Curve and Surface Fitting with Splines* (Oxford University, 1993).
6. P. Bo, R. Ling, and W. Wang, "A revisit to fitting parametric surfaces to point cloud," *Comput. Graph.* **36**(5), 534–540 (2012).
7. M. Ares and S. Royo, "Comparison of cubic B-spline and Zernike-fitting techniques in complex wavefront reconstruction," *Appl. Opt.* **45**(27), 6954–6964 (2006).
8. D. R. Iskander, M. J. Collins, and B. Davis, "Optimal modeling of corneal surfaces with Zernike polynomials," *IEEE Trans. Biomed. Eng.* **48**(1), 87–95 (2001).
9. M. K. Smolek and S. D. Klyce, "Zernike polynomial fitting fails to represent all visually significant corneal aberrations," *Invest. Ophthalmol. Vis. Sci.* **44**(11), 4676–4681 (2003).
10. V. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," *J. Opt. Soc. Am.* **71**(1), 75–85 (1981).
11. J. A. Díaz and R. Navarro, "Orthonormal polynomials for elliptical wavefronts with an arbitrary orientation," *Appl. Opt.* **53**(10), 2051–2057 (2014).
12. G. M. Dai, "Wavefront reconstruction from slope data within pupils of arbitrary shapes using iterative Fourier Transform," *Open Opt. J.* **1**(1), 1–3 (2007).
13. K. P. Murphy, *Machine Learning a Probabilistic Perspective* (Massachusetts Institute of Technology, 2012).
14. A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics* **12**(1), 55–67 (1970).
15. J. Schwiegerling, *Field Guide to Visual and Ophthalmic Optics* (SPIE, 2004).
16. M. M. Lipschutz, *Differential Geometry* (McGraw Hill, 1970).

17. D. Rodríguez-Ibañez, J. Alonso, and J. A. Quiroga, "Squareness error calibration of a CMM for quality control of ophthalmic lenses," *Int. J. Adv. Manuf. Technol.* **68**(1-4), 487–493 (2013).
18. J. Yao, J. Huang, P. Meemon, M. Ponting, and J. P. Rolland, "Simultaneous estimation of thickness and refractive index of layered gradient refractive index optics using a hybrid confocal-scan swept-source optical coherence tomography system," *Opt. Express* **23**(23), 30149–30164 (2015).
19. J. Wang, R. K. Leach, and X. Jiang, "Review of the mathematical foundations of data fusion techniques in surface metrology," *Surf. Topogr.: Metrol. Prop.* **3**(2), 023001 (2015).
20. J. A. Gómez-Pedrero, D. Rodríguez-Ibañez, J. Alonso, and J. A. Quiroga, "Design and development of a profilometer for the fast and accurate characterization of optical surfaces," *Proc. SPIE* **9628**, 96281I (2015).

1. Introduction

The robust and accurate fitting of a noisy cloud of 3D points to an analytical surface is a problem of paramount importance in fields such as computer aided design, virtual reality, computer vision and production engineering [1–3]. This problem was addressed first by Hayes and Hallyday [4] who presented a method for fitting a cloud of points to a B-Spline surface based on least-squares minimization of a functional defined as the Euclidean distance between the B-Splines control points and the measured cloud points. This technique has been further refined to address different situations such as scattered point clouds or point cloud smoothing [5]. Other authors have focused in enhancing the basic least-squares algorithm by considering different forms for the functional to be minimized. In the literature, the technique consisting in the minimization of the functional based on the Euclidean distance is known as the point distance minimization (PDM) method, while the method which consists in the minimization of an alternative functional based on the projected distance along the normal direction of the B-Spline surface at the control point is called the tangent distance method or TMD [1]. More sophisticated functionals which incorporate the information about the local curvatures have been proposed by Wang et al. [1] and Bo et al. [6]. Other refinements of the basic technique consist in the addition of different terms to the functional such as a regularizing term to avoid overfitting and a smoothness term in order to avoid abrupt variations in the local curvatures [6] at the surface borders.

It is worth to note that, for the techniques mentioned in the previous paragraph, the surface to be fitted to the cloud of points is a B-spline function. In optics, however, it is usual to fit cloud points to Zernike circle polynomials as they are defined in circular domains which match with the shape of many optical components such as lenses or pupils. We can find examples of this usage in the fitting of wavefront surfaces [7] but also in fitting data to physical surfaces like the surface of the cornea [8]. Despite this extensive usage, fitting a cloud of points to a set of Zernike polynomials presents also some drawbacks, as they are not able to describe complex wavefronts [7] and, also, as the standard fitting technique has shown limitations in the accuracy of the fitted surface [9]. Moreover, being polynomials defined in a circular domain they do not adjust well to other domain shapes such as annular or elliptical ones. For those cases, new families of orthogonal polynomials have been devised [10,11] but their analytical forms are quite complicated and they constitute particular cases difficult to generalize for complex regions.

In this work we present a new technique for fitting Zernike polynomials to noisy clouds of points defined over any connected region of arbitrary form located within the limits of the unit circle where the polynomials are defined. We make use of the formalism of machine learning which provides us with a more flexible way to deal with the fitting problem, particularly in defining a suitable cost functional for fitting free form ophthalmic surfaces, in order to obtain the desired accuracy for both surface elevation and local curvatures. Although the technique has been devised with this particular application in mind (fitting free-form ophthalmic lens surfaces) it can be readily adapted to other situations such as wavefront fitting where similar techniques have been published to get the wavefront from slope measurements [12]. In this latter case we can find a potential issue as it is well known that the Zernike circle polynomials are orthonormal only over circular domains. However, in our application we are more interested in describing a progressive surface as a sum of analytical functions, in this case Zernike polynomials (although the proposed technique can be readily adapted to other families of functions such as Chebyshev polynomials or Gaussian functions) and, in any case,

we could always perform a Gram-Smidt orthonormalization process to get an orthonormal family of polynomials over the domain of interest. Thus we have retained the expansion in Zernike polynomials as it constitutes one of the preferred ways of describing free form surfaces in optics.

The paper is organized as follows: in the next section we will show the theoretical basis of our technique. Afterwards, we will present the results obtained for both synthetic and real data and, finally, we will draw the relevant conclusions to finish the paper.

2. Theoretical foundations

Before presenting our technique we will give some brief definitions of the key concepts of the Machine Learning techniques employed to fit data to models, further information is given in reference [13] or any other generic book on Machine Learning. The goal of Machine Learning is to deduce automatically patterns from a set of data, which can be done with or without previous knowledge about the data. In the former case we speak of “supervised learning” while the latter case is known as “unsupervised learning”. In this context, data fitting is a special case of “supervised learning” as we have a set of input points which maps into another set of output points, being both sets known in advance, so that the goal of data fitting is the determination of the optimum values of the constants characterizing the function which models the mapping between the input and output set of points. In the language of Machine Learning we refer to the “training set” as the set formed by both input and output points. As we will see later, the usual way to get the coefficients of the fit is through the minimization of an error function [13] whose inputs are the points of the training set.

Therefore, in the context of Machine Learning [13] the problem of data fitting can be stated in the following terms. The set of data points can be represented by a collection of M row vectors called features so $\mathbf{x}^{(i)} = \{x_0^{(i)}, x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)}\}_{i=1,2,\dots,M}$, being customary to set $x_0^{(i)} = 1 \forall i$.

According to this definition, M is the number of points (or examples as they are known in the context of Machine Learning) of the data point set while N is the number of features considered which corresponds to the dimension of the feature vector $\mathbf{x}^{(i)}$. Together with the features, an output vector $\mathbf{y} = \{y^{(1)}, y^{(2)}, y^{(3)}, \dots, y^{(M)}\}_{i=1,2,\dots,M}$ is defined so that the set of pairs $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1,2,\dots,M}$ form the training set, as stated before. In order to relate the features with the output, a function known as hypothesis h is defined so that

$$y^{(i)} = h(\mathbf{x}^{(i)}). \quad (1)$$

In our case we will perform a linear optimization, so the hypothesis function will have the form

$$h(\mathbf{x}^{(i)}, \boldsymbol{\theta}) = \sum_{j=0}^N \theta_j x_j^{(i)} \equiv \mathbf{x}^{(i)} \cdot \boldsymbol{\theta}. \quad (2)$$

Where the elements of the column vector $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_N\}^T$ with T the transpose operator, are called the parameters or weights of the hypothesis function (or fitting coefficients according to the definition employed in the first paragraph of this section). In these conditions the goal of the optimization is to set the optimum values of the parameters by minimizing a cost functional. An example of this functional is the following

$$J(\boldsymbol{\theta}) = \frac{1}{2M} \sum_{i=1}^M \left(y^{(i)} - h(\mathbf{x}^{(i)}, \boldsymbol{\theta}) \right)^2 + \lambda \sum_{j=1}^N \theta_j, \quad (3)$$

Where the first term of the right side of Eq. (3) is the ordinary least squares term and the second one is known as the regularization term (being λ the regularization parameter) which is included to minimize the risk of overfitting. Equation (3) is referred in the literature as ridge

regression [13,14] and it can be solved using two approaches: an iterative one, known as gradient descent or a direct solution based in the so-called normal equation. In this second case, the solution of (3) is given by the following expression [13,14]

$$\boldsymbol{\theta}_{opt} = (\lambda \mathbf{I}_D + \mathbf{X}^T \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}^T, \quad (4)$$

where \mathbf{X} is a matrix defined from the feature vectors $\mathbf{x}^{(i)}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(M)} \end{bmatrix}, \quad (5)$$

and \mathbf{I}_D is a variation of the $(M \times N)$ identity matrix which takes into account the fact that the parameter θ_0 is not regularized so its mathematical form is:

$$\mathbf{I}_D = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (6)$$

Compared to the standard least squares fitting problem (which corresponds to the case $\lambda = 0$) Ridge regression provides a modified solution [14] which may present lower square error depending on the value of the regularization parameter λ . This means that if we denote $\boldsymbol{\theta}_{ls}$ as the solution of the standard least square fitting with $\lambda = 0$ and $\boldsymbol{\theta}_{opt}$ is the solution of the ridge regression given by Eq. (4) then

$$\boldsymbol{\theta}_{opt} = \boldsymbol{\theta}_{ls} - \lambda \mathbf{W} \boldsymbol{\theta}_{ls}, \quad (7)$$

where $\mathbf{W} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}$ according to reference [14]. Equation (7) summarizes the basic idea of ridge regression, that is, to substitute the standard least squares solution by a modified one which would reduce the errors between the experimental data and the linear model. However, it is necessary to take into account the fact that although there is always a value of λ with lower mean square errors, according to the existence theorem stated in reference [14]; for high values of λ , the solution given by ridge regression presents greater mean square error than the standard solution due to the appearance of a so called bias term [14]. Therefore, the selection of the correct value of the regression parameter λ is of paramount importance in the context of ridge regression, and we will explain latter the procedure that we have followed in order to get the optimum value of the regularization parameter λ which minimizes the effect of the bias in the mean square error introduced by the ridge regression.

We will adapt now the formalism of ridge regression which was stated in the preceding paragraphs to our particular problem of fitting a cloud of points to a surface defined as a combination of Zernike circle polynomials. In a Cartesian coordinate system, the cloud is given by a set of M points $\{\mathbf{r}^{(i)} = (\xi^{(i)}, \eta^{(i)}, z^{(i)})\}_{i=1,2,\dots,M} \in \mathbf{R}^3$ which are defined within a compact and connected domain $D \subset C(R)$, being $C(R)$ a circle of radius R in \mathbf{R}^2 centered at the origin of the coordinate system. In these conditions, the output vector is given by the set of elevations (the third coordinate of each measured point) so $\mathbf{y} = (z^{(1)}, z^{(2)}, \dots, z^{(M)})$. Our aim is to fit these data to a set of Zernike polynomials defined in a Cartesian coordinate as [15]

$$\begin{aligned}
& \text{For } m > 0 \\
Z_n^m(\xi, \eta) &= \sqrt{2(n+1)} \sum_{s=0}^{\frac{n-|m|}{2}} \sum_{j=0}^{\frac{n-|m|}{2}-s} \sum_{k=0}^{\frac{n-|m|}{2}-s-j} \frac{(-1)^{s+k} (n-s)!}{s! \left[\frac{n+|m|}{2} - s \right]! \left[\frac{n+|m|}{2} - s - k \right]!} \binom{\frac{n-m}{2} - s}{j} \binom{m}{2k} \xi^{n-2(s+j+k)} \eta^{2(j+k)} \\
& \text{For } m < 0 \\
Z_n^m(\xi, \eta) &= \sqrt{2(n+1)} \sum_{s=0}^{\frac{n-|m|}{2}} \sum_{j=0}^{\frac{n-|m|}{2}-s} \sum_{k=0}^{\frac{n-|m|}{2}-s-j} \frac{(-1)^{s+k} (n-s)!}{s! \left[\frac{n+|m|}{2} - s \right]! \left[\frac{n+|m|}{2} - s - k \right]!} \binom{\frac{n-m}{2} - s}{j} \binom{m}{2k+1} \xi^{n-2(s+j+k)} \eta^{2(j+k)+1} \\
& \text{For } m = 0 \\
Z_n^0(\xi, \eta) &= \sqrt{n+1} \sum_{s=0}^{n/2} \sum_{j=0}^{n/2-s} \frac{(-1)^s (n-s)!}{s! \left(\frac{n}{2} - s \right)! \left(\frac{n}{2} - s \right)!} \binom{\frac{n}{2} - s}{j} \xi^{n-2(s+j)} \eta^{2j}.
\end{aligned} \tag{8}$$

Note that, in Eq. (8) n and m stands for the indices of the Zernike polynomial, but in this work we will adopt instead a single indexing scheme as the one proposed in reference [15]. To do so, we will define the set of polynomials as $Z_j(\xi, \eta)_{j=0,1,\dots,N}$ being $j = \frac{n(n+2)+m}{2}$. As the Zernike polynomials given by Eq. (8) are normalized for the unit circle, it is necessary to perform a normalization of the coordinates $(\xi^{(i)}, \eta^{(i)})_{i=1,2,\dots,M}$ so we will define the set normalized coordinates as

$$\hat{P}^{(i)} = \left(\frac{\xi^{(i)}}{R}, \frac{\eta^{(i)}}{R} \right), \tag{9}$$

being R the radius of the circle which contains the domain D where the cloud is defined, as stated before. In these conditions, we will define the feature vector as the set of values resulting from the evaluation of each Zernike polynomial at a given point, that is

$$\mathbf{x}^{(i)} = \left(Z_0(\hat{P}^{(i)}), Z_1(\hat{P}^{(i)}), Z_2(\hat{P}^{(i)}), \dots, Z_N(\hat{P}^{(i)}) \right)_{i=1,2,\dots,M}, \tag{10}$$

so the hypothesis function will be in our case

$$h(\mathbf{x}^{(i)}, \boldsymbol{\theta}) = \theta_0 Z_0(\hat{P}^{(i)}) + \theta_1 Z_1(\hat{P}^{(i)}) + \dots + \theta_N Z_N(\hat{P}^{(i)}), \tag{11}$$

and the parameters of our model, given by vector $\boldsymbol{\theta}$, are the coefficients of the series of Zernike polynomials which best fit our data. These coefficients would be obtained through Eq. (4). Note that, $Z_0(\hat{P}^{(i)}) = 1 \forall \hat{P}^{(i)}$ as this polynomial corresponds to the piston term $Z_0^0(\xi, \eta) = 1 \forall (\xi, \eta)$ of the Zernike expansion (8).

In our particular problem of fitting a free-form surface corresponding to an ophthalmic lens, it is not enough that the elevation of the fitted surface corresponds accurately with the elevation of the points of the cloud, it is also extremely important to ensure the accuracy of the local curvatures of the fitted surface, because any error in those curvatures would automatically be translated into an error in the lens power. In particular, we are interested in the main curvatures κ_1 and κ_2 which for a surface defined as a Monge's form are directly related to the Gaussian and median curvature as [16]

$$\begin{aligned}\kappa_1 &= H - \sqrt{H^2 - K}, \\ \kappa_2 &= H + \sqrt{H^2 - K}.\end{aligned}\quad (12)$$

Where H is the mean and K is the Gaussian curvature which are related to the derivatives of the elevation $s(x, y)$ of the surface when it is described as a Monge's form so that [16]

$$\begin{aligned}H &= \frac{\left(1 + \left(\frac{\partial s}{\partial x}\right)^2\right) \frac{\partial^2 s}{\partial x^2} - 2 \frac{\partial s}{\partial x} \frac{\partial s}{\partial y} \frac{\partial^2 s}{\partial x \partial y} + \left(1 + \left(\frac{\partial s}{\partial y}\right)^2\right) \frac{\partial^2 s}{\partial y^2}}{\left(1 + \left(\frac{\partial s}{\partial x}\right)^2 + \left(\frac{\partial s}{\partial y}\right)^2\right)^{3/2}}, \\ K &= \frac{\frac{\partial^2 s}{\partial x^2} \frac{\partial^2 s}{\partial y^2} - \left(\frac{\partial^2 s}{\partial x \partial y}\right)^2}{1 + \left(\frac{\partial s}{\partial x}\right)^2 + \left(\frac{\partial s}{\partial y}\right)^2}.\end{aligned}\quad (13)$$

In these conditions, in order to avoid the errors associated to the local curvatures two mathematical conditions must be accomplished: First, the local curvatures are expected to be continuous and differentiable, which is guaranteed as long as we describe the surface as a polynomial series. Secondly, they are also assumed to change slowly along the lens surface (brusque curvature variations could be unpleasant and even annoying for the lens user so they would impair the lens performance). Mathematically, we will express this latter condition as follows

$$\begin{aligned}\nabla \left(\frac{\partial^2 s(x, y)}{\partial x^2} \right) &= \mathbf{0}, \\ \nabla \left(\frac{\partial^2 s(x, y)}{\partial y^2} \right) &= \mathbf{0},\end{aligned}\quad (14)$$

where ∇ stands for the gradient operator. Equation (14) implies that the gradient of the second derivatives of surface elevation, which are related to the local curvatures through Eqs. (12) and (13), is close zero which guarantees a slow variation of the local curvatures. We will include this condition in our model through an extension of the training set showing in this way the flexibility of the machine learning formalism to be easily adapted to different cases. To do so, we will define first an additional set of points defined over the unit circle as

$$\hat{P}^{(i)} = \left(\frac{\xi^{(i)}}{R}, \frac{\eta^{(i)}}{R} \right)_{i=1,2,\dots,M'}.\quad (15)$$

Note that the original set of points $\{\hat{P}^{(i)}\}$ and the new set $\{\hat{P}^{(i)}\}$ are different and some points of the new set $\{\hat{P}^{(i)}\}$ are even located outside the dominion D in which the cloud points to be fitted is defined (although they are always within the unit circle). This is important in order to ensure a smooth curvature variation even in the points of the surface close to the border of dominion D . In these conditions, we will define four new feature sets as

$$\begin{aligned}
\mathbf{x}^{(i)} &= \left(\frac{\partial^3 Z_0(\hat{P}^{(i)})}{\partial x^3}, \frac{\partial^3 Z_1(\hat{P}^{(i)})}{\partial x^3}, \dots, \frac{\partial^3 Z_N(\hat{P}^{(i)})}{\partial x^3} \right)_{i=1,2,\dots,M'}, \\
\mathbf{x}^{n(i)} &= \left(\frac{\partial^3 Z_0(\hat{P}^{(i)})}{\partial x \partial y^2}, \frac{\partial^3 Z_1(\hat{P}^{(i)})}{\partial x \partial y^2}, \dots, \frac{\partial^3 Z_N(\hat{P}^{(i)})}{\partial x \partial y^2} \right)_{i=1,2,\dots,M'}, \\
\mathbf{x}^{m(i)} &= \left(\frac{\partial^3 Z_0(\hat{P}^{(i)})}{\partial y \partial x^2}, \frac{\partial^3 Z_1(\hat{P}^{(i)})}{\partial y \partial x^2}, \dots, \frac{\partial^3 Z_N(\hat{P}^{(i)})}{\partial y \partial x^2} \right)_{i=1,2,\dots,M'}, \\
\mathbf{x}^{mm(i)} &= \left(\frac{\partial^3 Z_0(\hat{P}^{(i)})}{\partial y^3}, \frac{\partial^3 Z_1(\hat{P}^{(i)})}{\partial y^3}, \dots, \frac{\partial^3 Z_N(\hat{P}^{(i)})}{\partial y^3} \right)_{i=1,2,\dots,M'},
\end{aligned} \tag{16}$$

which will be combined with the original feature set to form the extended feature set defined as

$$\begin{aligned}
X_{ext} = \{ & \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}, \mu \mathbf{x}^{n(1)}, \dots, \mu \mathbf{x}^{n(M)}, \mu \mathbf{x}^{m(1)}, \dots, \mu \mathbf{x}^{m(M)}, \\
& \mu \mathbf{x}^{mm(1)}, \dots, \mu \mathbf{x}^{mm(M)} \},
\end{aligned} \tag{17}$$

with μ a parameter which modulates the relative importance assigned to the constant curvature in comparison with the sagitta errors. In the same way we have extended the training set the output vector must be also extended as follows

$$\mathbf{y}_{ext} = [\mathbf{y} \quad \mathbf{0}_{1 \times M'} \quad \mathbf{0}_{1 \times M'} \quad \mathbf{0}_{1 \times M'} \quad \mathbf{0}_{1 \times M'}], \tag{18}$$

where $\mathbf{0}_{1 \times M'}$ is a row vector of dimensions $(1 \times M')$, with all of their elements equal to zero. In these conditions if we define matrix \mathbf{X}_{ext} as

$$\mathbf{X}_{ext}(\mu) = \begin{bmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(M)} \\ \mu \mathbf{x}^{n(1)} \\ \vdots \\ \mu \mathbf{x}^{n(M)} \\ \mu \mathbf{x}^{m(1)} \\ \vdots \\ \mu \mathbf{x}^{m(M)} \\ \mu \mathbf{x}^{mm(1)} \\ \vdots \\ \mu \mathbf{x}^{mm(M)} \end{bmatrix}, \tag{19}$$

then we can compute the optimum values for the Zernike coefficients as

$$\boldsymbol{\theta}_{opt}(\lambda, \mu) = (\lambda \mathbf{I}_D + \mathbf{X}_{ext}^T(\mu) \cdot \mathbf{X}_{ext}(\mu))^{-1} \cdot \mathbf{X}_{ext}^T(\mu) \mathbf{y}_{ext}^T, \quad (20)$$

so the optimum values of the Zernike coefficients will depend on the election of the two parameters λ and μ , which will control the weight of the high order polynomials and the smoothness of the main curvatures of the fitting surface, respectively. In next section we will show the results obtained after applying our algorithm to simulated and measured surfaces.

3. Experimental results

We have carried out a number of experiments to test our model. In the first experiment we have fitted a synthetic data set. To do so we have generated first a progressive surface using a bi-variate polynomial of order 15 within the unit circle. Afterwards, we have added Gaussian noise with a standard deviation of $3 \cdot 10^{-4}$ mm and, finally, we have used a polygonal mask to define the surface's arbitrary domain within the unit circle. The final noisy surface is depicted in Fig. 1).

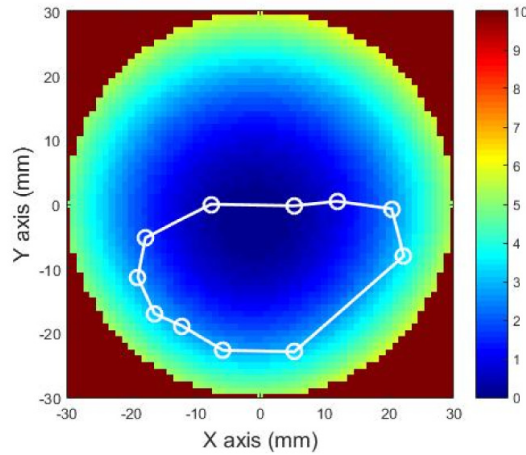


Fig. 1. Map of the elevation of the theoretical surface generated together with the polygonal path (white line with circular dots) which defines the region of interest (ROI) where the cloud of points is defined.

We have fitted this surface to the set composed by the first 209 Zernike polynomials. Note that the number of degrees of freedom of the synthetic surface is around 120 which corresponds to the number of monomials of the surface, but we have chosen a higher number of terms for the Zernike expansion. Although, for the surface selected this may cause overfitting problems (which may be alleviated by the regularization term as we will see latter) in a realistic case, we would not know in advance the complexity of the surface, so we would need a high number of polynomials so that we can measure any kind of surface regardless of its complexity. In order to get the residual errors, we have fitted the whole surface, i.e. that which is defined over the circular domain, to the Zernike polynomials and we have done the same for the subset defined over the ROI shown in Fig. 1). In both cases, in order to get the residual error, we have not added noise to the original surface and we have set $\lambda = 0$ and $\mu = 0$, so neither regularization nor curvature smoothness were employed. We have computed the elevation error which is defined as the standard deviation of the distribution of the residues between the elevations of the fitted surface and that of the theoretical one. In these conditions, the elevation error obtained is $\delta z = 1.4 \cdot 10^{-5}$ mm for the whole surface and $\delta z = 4.8 \cdot 10^{-5}$ mm for the subset surface. The lower elevation error obtained for the whole surface is due to the fact that Zernike polynomials are orthogonal within the unit circle and that we have not used the regularization terms. Regarding the values of the expansion coefficients obtained, we have depicted in Fig. 2(a) the coefficients obtained for the Zernike series over the whole surface

and in Fig. 2(b) the coefficients obtained when fitting the subset surface defined over the non-regular domain. As it can be expected [12], there are differences between these two sets of coefficients as the surface which they are fitting are defined over different domains. As it should be expected and it is shown in Fig. 2(b) the standard fit of Zernike polynomials over a cloud of points defined in a non-circular region clearly presents an overfitting with huge variations of the Zernike coefficients.

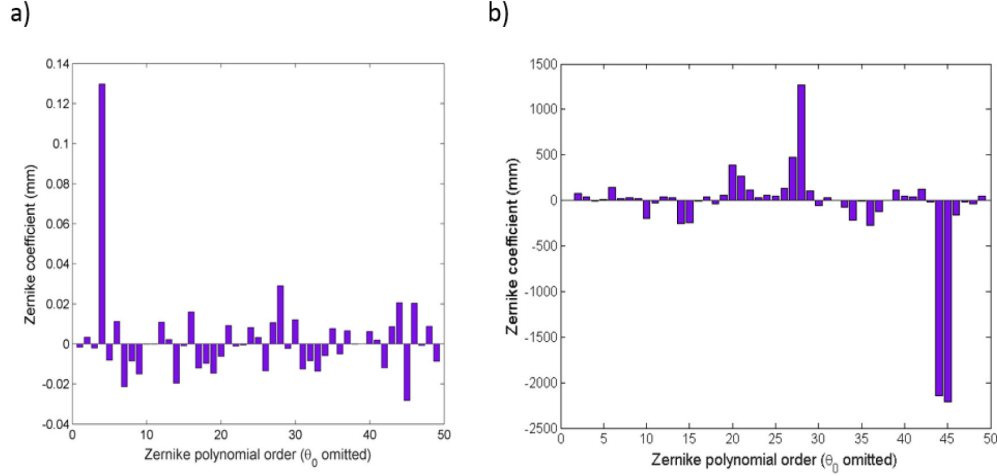


Fig. 2. a) Plot of the values of the first 50 coefficients corresponding to the standard fit (with $\lambda = 0$ and $\mu = 0$) of a Zernike polynomial series to the theoretical surface of Fig. 1), defined over the entire circular region, b) values of the coefficients corresponding to the fit of the same surface but defined over the ROI delimited by the white line of Fig. 1). In both plots the piston term have been removed for clarity.

After performing the standard fitting, we have fitted the surface defined over the irregular domain using two different training sets, those defined in Eqs. (10) and (17) with the cost function given in Eq. (3). In order to obtain the optimum values of the parameters λ and μ , we have followed the next procedure: First, we have varied the value of the regularization parameter λ keeping $\mu = 0$ and we have computed the elevation error δz within the ROI for each value of λ . We have done this calculation adding Gaussian noise to the theoretical surface with three different values of the standard deviation, 10^{-4} , $3 \cdot 10^{-4}$ and 10^{-3} mm, being the result shown in Fig. 3(a).

As it can be seen in Fig. 3(a), regardless of the amount of noise, the elevation error varies with λ in a similar fashion: First, the elevation error decreases with λ and then, after reaching a minimum, it raises due to the effect of the bias associated to the ridge regression [14]. As we can see in Fig. 3(a), the optimum value of λ increases with the noise, so that we have chosen the value $\lambda = 10^{-6}$ which corresponds to the minimum of the elevation error for the noise with standard deviation of $3 \cdot 10^{-4}$ mm. We have selected this value of noise because, as we will see later, it matches with the accuracy of the profilometer that we have employed for measuring the surfaces of the lenses. Once we have obtained the optimum value for λ , we have followed the same procedure to obtain μ , that is, we have varied the value of μ (keeping $\lambda = 10^{-6}$) and we have computed the corresponding elevation error for three different noise levels. As it can be seen in Fig. 3(b), for the intermediate noise with standard deviation of $3 \cdot 10^{-4}$ mm, the optimum value of μ is around $3 \cdot 10^{-5}$.

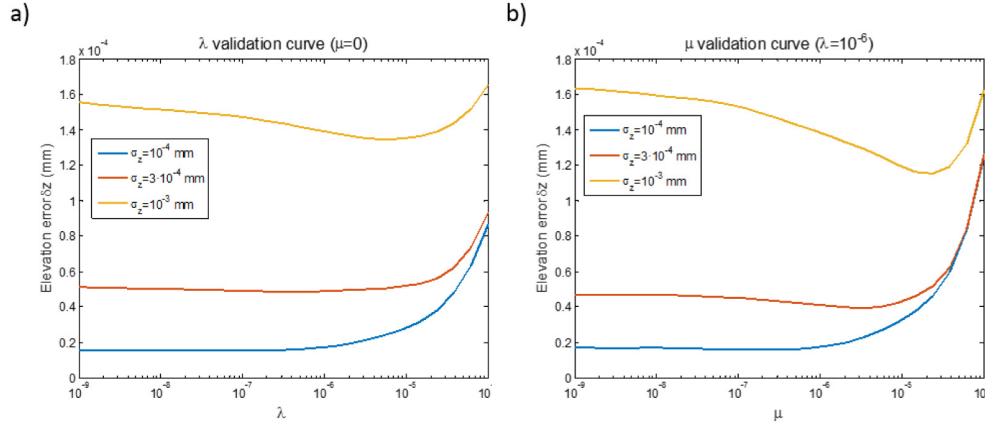


Fig. 3. a) Plot of the elevation error versus the regularization parameter λ for three different levels of noise added to the theoretical surface, b) plot of the elevation error versus the curvature smoothing parameter μ for the same noise levels and $\lambda = 10^{-6}$. These plots are useful in order to get the optimum values of the parameters λ and μ .

Therefore we have performed the fit for the theoretical surface selecting for the first training set the value of the regularization parameter of $\lambda = 10^{-6}$, while for the extended training set we kept the same value for the regularization parameter and we fixed $\mu = 3 \cdot 10^{-5}$, and we have added to the theoretical surface a noise with Gaussian deviation of $3 \cdot 10^{-4}$ mm. In these conditions, we have computed the error maps that can be seen in Fig. 4). The analysis of the fitting residuals, computed in a region within the ROI in order to avoid distortions by the border points results in a mean value of $0.1 \mu\text{m}$ and a standard deviation of $0.08 \mu\text{m}$ for the first training set, see Fig. 4(a), and a mean of $0.08 \mu\text{m}$ with a standard deviation of $0.06 \mu\text{m}$ for the extended training set as it can be seen in Fig. 4(b). Therefore, in spite of the added noise, the algorithm has been able to fit the surfaces keeping the elevation errors below 1 micron according to the standard fixed by the ophthalmic lens manufacturing industry [17].

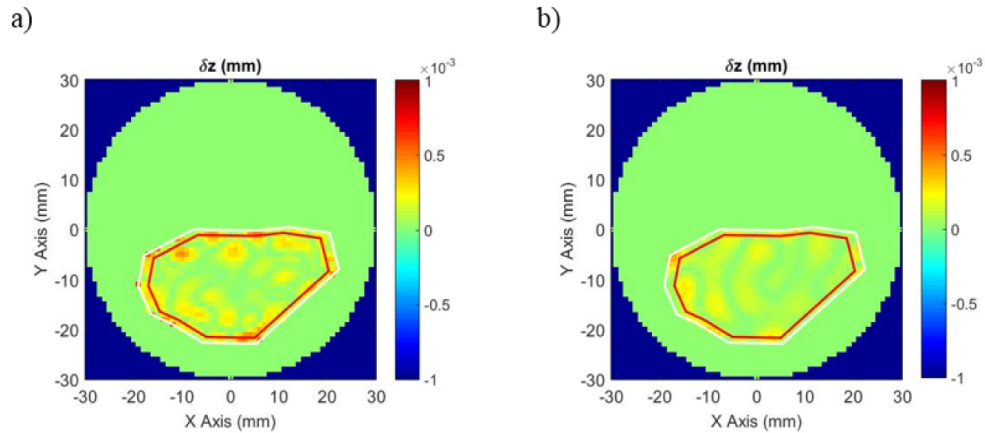


Fig. 4. a) Map of the elevation error along the lens surface obtained for the first training set, i.e., setting $\lambda = 10^{-6}$ and $\mu = 0$ and b) elevation error map obtained for the extended training set with $\lambda = 10^{-6}$ and $\mu = 3 \cdot 10^{-5}$. In both cases the elevation error is lower than 1 micron for the ROI. The ROI is limited by the white line and the red line delimits an inner region for computing the statistical magnitudes (mean and standard deviation) of the fitting residues. The green area outside the ROI represents the circular domain where the Zernike polynomials are defined.

However, as stated before, the assessment of the elevation error is not enough in our case, because we must check also that brusque variations in surface curvature are not present so that the surface can be used in an ophthalmic lens. To do so, we have first computed the main curvatures κ_1 and κ_2 , being always κ_2 the greater one, for both the fitted and theoretical surface. From these curvatures, we have defined the local sphere as

$$S = (n-1)(\kappa_2 + \kappa_1) / 2, \quad (21)$$

and the local cylinder as

$$C = (n-1)(\kappa_2 - \kappa_1), \quad (22)$$

being n the refractive index which have been taken as $n = 1.55$ in our computations. We use the local sphere and cylinder because if the surface forms part of an ophthalmic lens, they are directly related with the sphere and cylinder experienced by the lens user. In these conditions, the lens manufacturers use to characterize the surface errors in terms of local sphere and cylindrical errors [17]. In fact, the usual rule followed by the industry is to keep the errors in local sphere and cylinder lower than 0.1 D (ideally, the value should be close to 0.01 D) along the whole surface while the elevation error should be lower than $1 \mu\text{m}$ for any point of the surface.

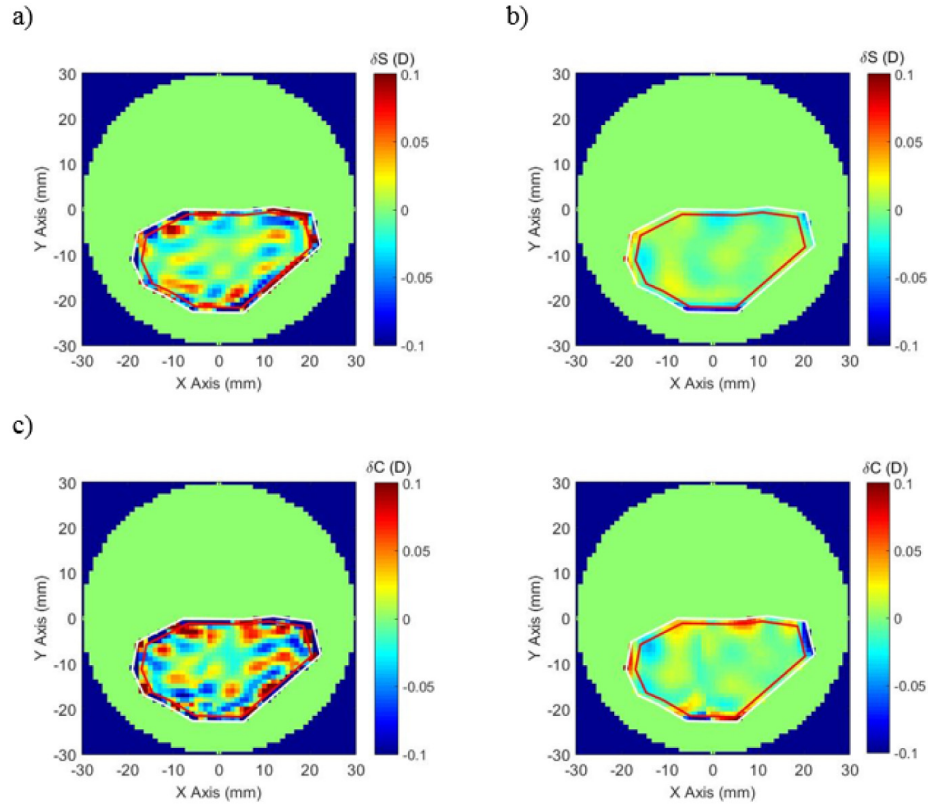


Fig. 5. a) Map of the spherical error in the ROI obtained for a theoretical progressive surface by setting $\lambda = 10^{-6}$ and $\mu = 0$ (original training set), b) map of the equivalent spherical error for the same surface obtained using the extended training set ($\lambda = 10^{-6}$ and $\mu = 3 \cdot 10^{-5}$), c) map of the cylindrical error obtained with the original training set and d) map of the cylindrical error obtained with the extended training set. The effect of the training set extension can be noticed as a reduction on the errors thorough the whole ROI particularly for the spherical equivalent power.

In Fig. 5) we have plotted the equivalent sphere and cylinder error maps for the two training sets employed. We have also computed the mean and standard deviation for these errors within a region located inside of the surface domain (marked with red line in the figures as in the preceding case). As it can be seen in Fig. 5(a) the error in the equivalent spherical refractive power obtained for the first training set is within the tolerance of 0.1 D and the same happens with the cylindrical power depicted in Fig. 5(c). However, we can appreciate an increment in the errors for the points located close to the border of the ROI selected. The high values of spherical equivalent and cylindrical at the ROI borders can be lowered if we enforce the condition of smooth curvature variation by setting a value of $\mu = 3 \cdot 10^{-5}$ as it can be appreciated in the maps depicted in Figs. 5(b) and 5(d). These tendencies are confirmed by the numerical results. For the spherical equivalent power error, the mean error is 0.02 D and the standard deviation is 0.03 D for the original training set. For this set, the cylindrical power error has mean 0.03 D with a standard deviation of 0.04 D. If the extended set is employed, the value of the mean is lowered to 0.007 D for the spherical equivalent power error and 0.01 D for the cylindrical power error, being the values of the standard deviation 0.009 D for the spherical equivalent and 0.013 for the cylinder. Therefore, we can conclude that using the extended training set a reduction on the standard deviation of the spherical and cylindrical errors is achieved.

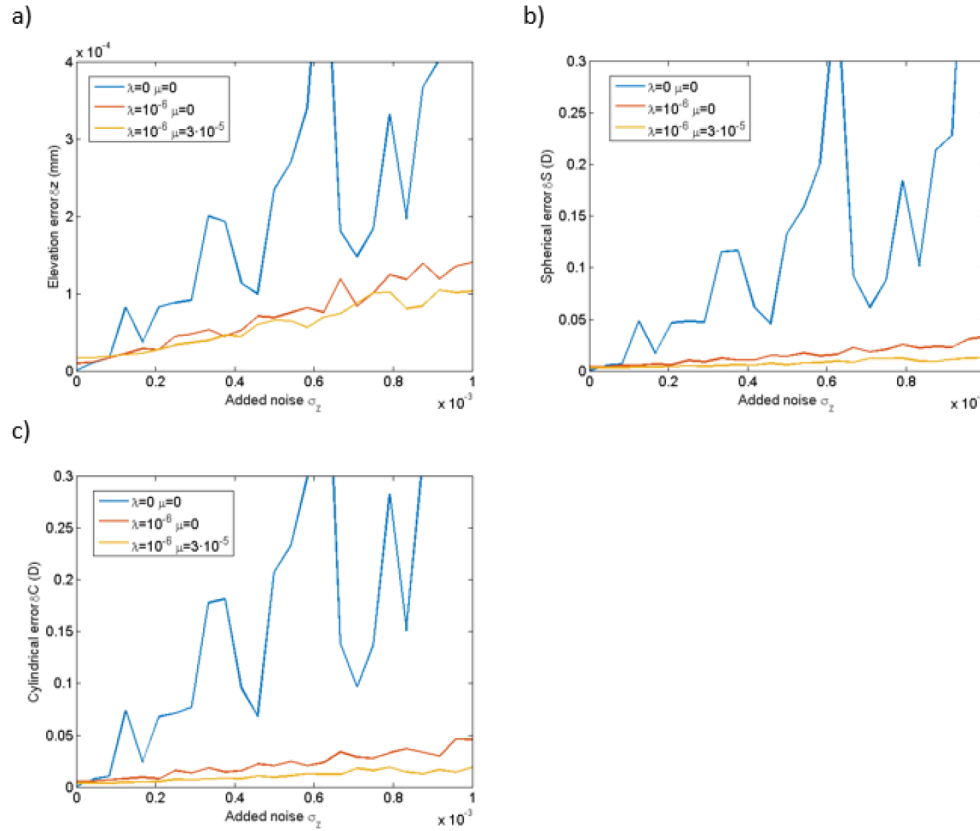


Fig. 6. a) Plot of the elevation error recovered for different values of the standard deviation of the added Gaussian noise, b) plot of the spherical error and c) cylindrical error versus Gaussian noise. We have computed the value of error for three cases: no regularization and no curvature smoothing (blue curve), regularization but no curvature smoothing (red curve) and regularization and curvature smoothing (yellow curve).

In order to state more clearly the accuracy of the fitting technique proposed we have followed a procedure, similar to that described in references [18,19], which consists in the

computation of the elevation, sphere and cylinder errors for the theoretical surface within the same ROI with different values of the added noise Gaussian. We have computed three different cases: The first one is the standard fit with no regularization neither curvature smoothing, that is $\lambda = 0$ and $\mu = 0$. For the second case, $\lambda = 10^{-6}$ and $\mu = 0$, that is we have considered regularization but no curvature smoothing and, finally, we have set the values of parameters λ and μ to the optimum ones as determined from the plots of Fig. 3), so that we have fixed $\lambda = 10^{-6}$ and $\mu = 3 \cdot 10^{-5}$. In Fig. 6) we have plotted the results obtained and we can see how the regularization term lowers the error in both the elevation and the spherical and cylindrical powers and how the introduction of a curvature smoothing diminishes the errors in the local curvatures as it can be appreciated in Figs. 6(b) and 6(c). Regarding the magnitude of the errors, although we have observed certain dependence with the ROI size and shape we can conclude that both the elevation and curvature errors are below the limits fixed by the ophthalmic industry (1 micron for elevation and 0.1 D for power).

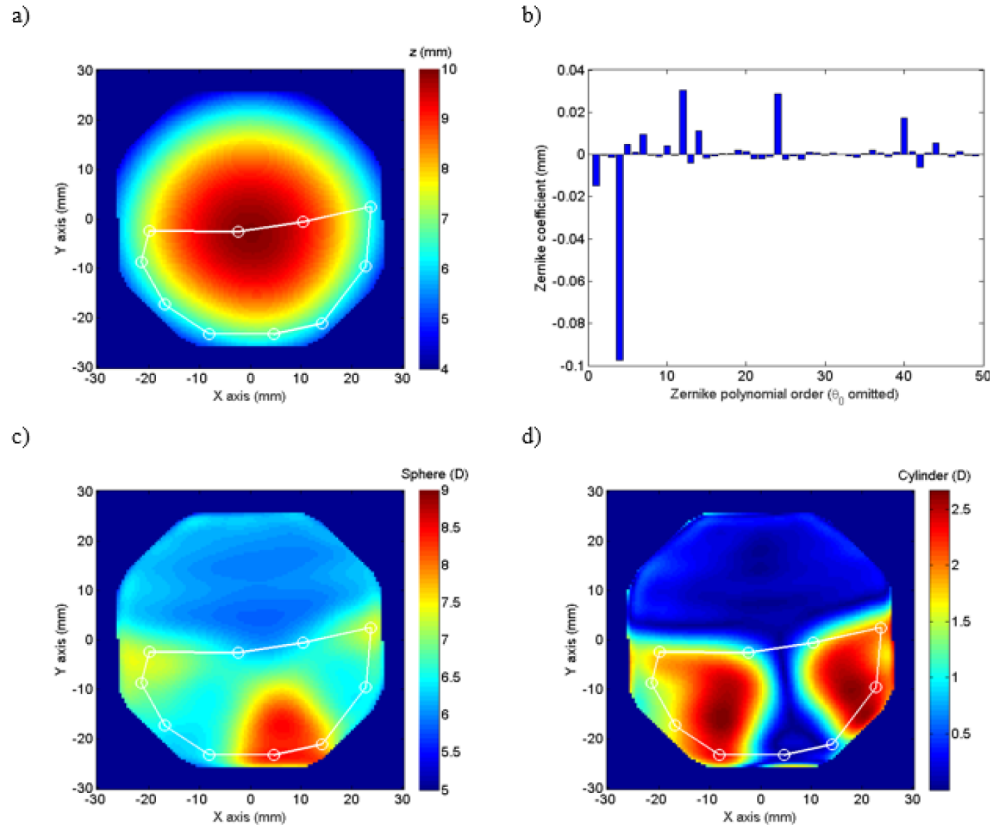


Fig. 7. a) Elevation map measured with the CMM of a progressive surface. b) Plot of the 50 first coefficients of the Zernike polynomial expansion corresponding to the surface defined in Fig. 7(a) for the whole circular domain. c) Equivalent sphere and b) cylindrical refractive power maps computed for the reference surface from the elevation map of Fig. 7(a). As in former cases, we have defined a ROI region whose border is given by the white line with circular markers.

In a second experiment we have measured the progressive surface of a progressive power lens using a coordinate measuring machine (CMM) [17,20]. The CMM used is a custom made one, with a resolution of 10^{-4} mm in the X, Y and Z directions [20] and an accuracy estimated in around $3 \cdot 10^{-4}$ mm which corresponds to the standard deviation of the residues obtained after measuring a reference spherical surface [20]. As the lens surfaces are defined within a

circle of 25 mm of radius, we have computed the Zernike polynomial series for an arbitrary region within the circle defining the lens. For comparison purposes, we have computed the spherical equivalent and cylindrical refractive power maps for the whole surface using a standard Zernike fitting in a circular domain, see Fig. 7) and we have compared these maps in the ROI with the ones obtained after fitting the surface using our technique. In this way we will be able to show how our technique is able to fit a surface within an arbitrary region obtaining the same power maps (within this region) as the ones resulting from a conventional Zernike polynomial fitting in a circular domain. We have also computed the elevation error defined as the difference between the elevation of the points measured within the ROI and the elevation of the points computed from the Zernike fit using our technique.

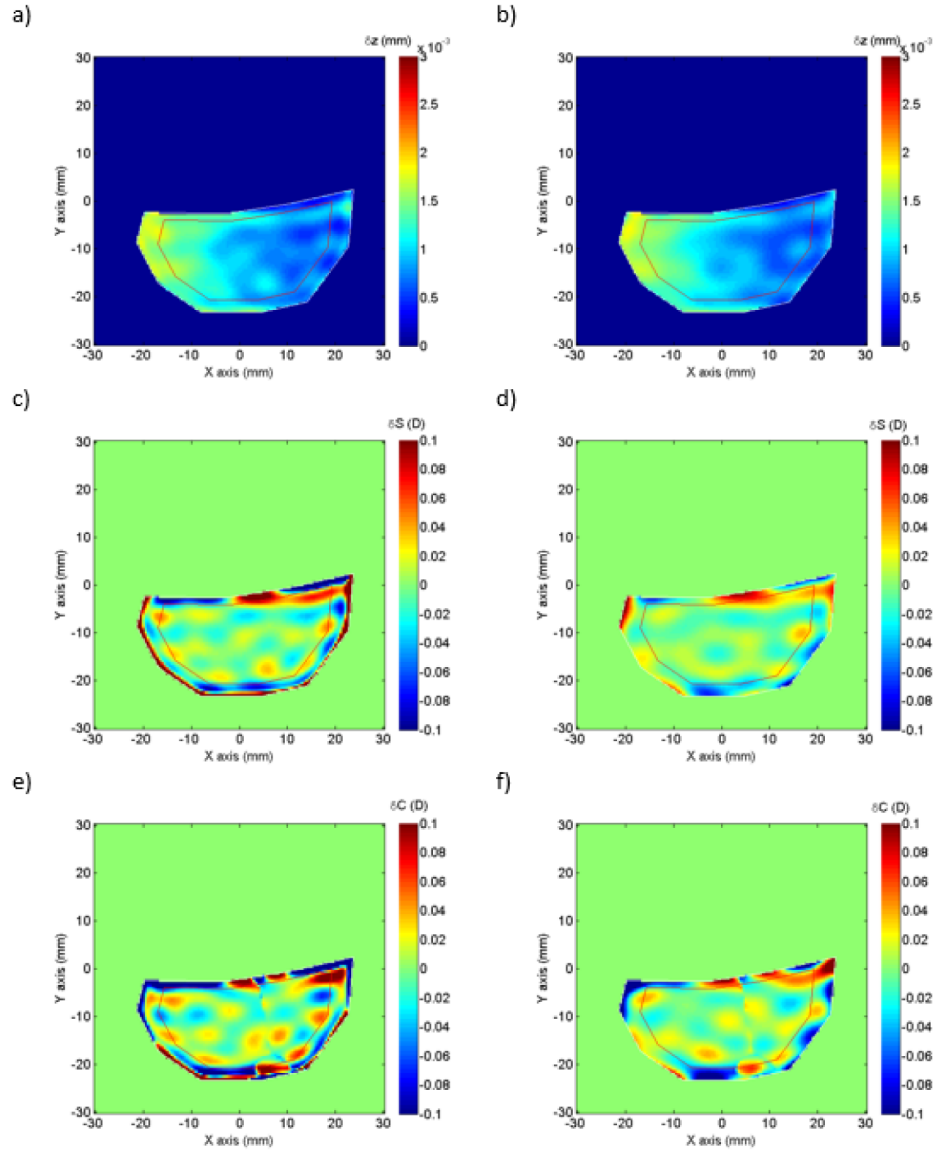


Fig. 8. a) Distribution of the elevation error without curvature smoothing and b) with curvature smoothing, c) difference of the equivalent spherical refractive power between the measured and reconstructed surfaces with no curvature smoothing, d) Same difference as in case c) but with curvature smoothing. Difference between the measured and reconstructed cylindrical refractive power e) with no curvature smoothing and f) with curvature smoothing.

In Fig. 8) we have represented the elevation error as defined in the preceeding paragraph, together with the difference between spherical equivalent and cylindrical refractive power maps computed using our technique and the conventional Zernike fitting to a circular domain. The measured power maps have been computed with and without curvature smoothing. For the first case, the fit parameters have been set as $\lambda = 10^{-6}$, $\mu = 0$, while for the second case we have set $\lambda = 10^{-6}$, $\mu = 3 \cdot 10^{-5}$ in order to force a smoother variation of the local curvatures.

As it can be seen in Figs. 8(a) and 8(b) the elevation error is kept below 1 micron for the ROI considered with an standard deviation of $2.7 \cdot 10^{-4}$ mm with no curvature smoothing and $2.6 \cdot 10^{-4}$ mm with curvature smoothing which are quite similar as expected as the curvature smoothing has no effect on the elevation values as seen in Fig. 6(a). In Figs. 8(c)-8(f) we can see that the errors are lower for both the spherical and cylindrical refractive powers, when the Zernike expansion is computed with $\mu \neq 0$ that is, when the smoothness of local curvatures is enforced. This is confirmed by the values of the standard deviation of the differences between the measured and original power maps.

With no curvature smoothing, the standard deviation is 0.02 D for the spherical power error and 0.025 D for the cylindrical error. When the extended training set is chosen, then the standard deviation of the mean spherical equivalent error is 0.014 D and the standard deviation for the cylindrical error is 0.017 D. These values are greater than the ones obtained for the fitting to the theoretical surface due, basically, to the measurement noise and to the strong curvature variation which present the surface in the ROI selected as can be appreciated in Figs. 7(c) and 7(d). In any case, the values of the standard deviation in both spherical equivalent and cylindrical error improve when the extended training set is selected.

The effect of the extended training set can be illustrated by representing in Fig. 9) the coefficients of the 50 first Zernike polynomials of the serial expansion with and without curvature smoothing. As it can be seen in Fig. 9(b) when we select the curvature smoothing fit, only Zernike polynomials with relatively lower index contribute to the fitting so the resulting surface will present smooth curvatures as it is composed by a sum of lower order Zernikes.

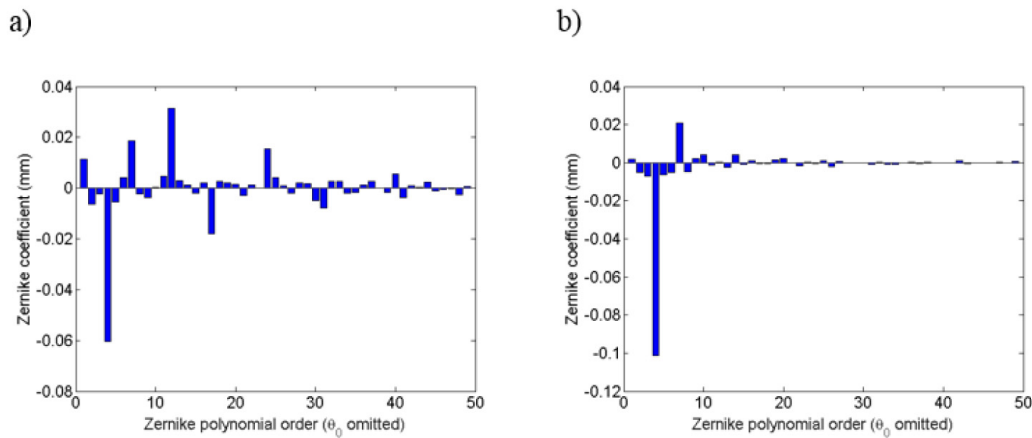


Fig. 9. Plot of the values of the first 50 coefficients corresponding to the fit of the surface to a Zernike polynomial series to the measured surface within the non regular region delimited by the white line in Figs. 7) and 8). The plot of panel a) correspond to the fitting with no curvature smoothing while that of panel b) corresponds to the fitting with curvature smoothing. In both plots the piston term have been removed for clarity.

4. Conclusions

A new algorithm for fitting a cloud of points defined over non regular region to a set of Zernike polynomials have been defined. The proposed algorithm uses the techniques of

machine learning, particularly an extension of the ridge regression in order to ensure a smooth variation of the surface curvatures through the ROI selected. In this way, the algorithm can be employed for fitting surfaces presenting smooth curvature variation as it happened with the ophthalmic lens surfaces, particularly for progressive surfaces. Despite having been developed for fitting ophthalmic lens surfaces, the algorithm can be readily adapted to other applications of interest in Optics, such fitting of wavefronts defined over non regular regions (non circular pupils).

The algorithm has been tested with both synthetic and real data. For synthetic data, we have shown that the elevation errors are kept below one micron and the curvature errors are lower than 0.1 D in agreement with the tolerances usually employed in the ophthalmic manufacturing industry.

We have also fitted a measured cloud of points corresponding to a surface measured using a CMM. In this case, we have compared our technique with a conventional Zernike fitting algorithm showing that, within the region of interest the errors in both elevation ($2.7 \cdot 10^{-4}$ mm) and spherical and cylindrical powers (0.015 D) are also under the tolerance value. We have also shown that using an extended training set enforces the condition of smooth curvature variation by selecting low order Zernike polynomials avoiding, in this way, overfitting problems.

Acknowledgments

The authors wish to thank to the Spanish Ministry of Economy and Competitiveness for the economic support provided under the grant DPI2012-36103.